SYMPOSIUM: DRUG RESIDUE AVOIDANCE: THE ISSUE OF TESTING

Quality Milk and Tests for Antibiotic Residues

WILLIAM M. SISCHO
Department of Veterinary Science,
The Pennsylvania State University, University Park 16802

ABSTRACT

One goal of total quality management is to prevent the occurrence of antibiotics in raw milk shipped from the farm. An effective approach to meet this goal is the implementation of HACCP (Hazard Analysis Critical Control Point) procedures, which are part of the Milk and Dairy Beef Quality Assurance Program for antibiotic avoidance. The program defines 10 critical control points, including screening tests for preventing antibiotic residues. Although milk from individual cows clearly should be tested to ensure that antibiotic-free milk is leaving the farm, it is not clear whether any existing tests can be reliably used on milk samples from individual cows, or even on samples from bulk tanks. The FDA acceptance procedures have not required that bulk milk tests undergo a population evaluation; these tests have not been objectively evaluated for individual cows. Of more concern, detection limits differ among tests, sometimes approaching zero. Despite the intent of the Pasteurized Milk Ordinance, milk acceptability definitions vary between states. In addition, the predictive value of test results has not been integrated into the regulatory process. Although largely ignored by the regulatory agencies, these issues cannot be ignored by the dairy industry. Ultimately, the milk testing program should become a component of the quality process that is centered on the farm and that measures the success of the industry in producing high quality milk rather than being a regulatory program that searches for a flawed product.
(**Key words**: food safety, milk quality, antibiotic residue tests)

Abbreviation key: **GAO** = General Accounting Office, **HACCP** = Hazard Analysis Critical Control Point, **MDBQAP** = Milk and Dairy Beef Quality Assurance Program, **PMO** = Pasteurized Milk Ordinance.

Received June 25, 1995.
Accepted December 14, 1995.

INTRODUCTION

Recent information shows that dairy producers are producing a high quality milk supply that has a very low prevalence of β-lactam antibiotic residues. In the US between October 1, 1993 and September 30, 1994, a total of 3,213,220 samples of bulk milk were tested at the time of tanker pickup; 2024 of these samples—representing approximately 28,500,000 kg of milk—were positive (1). This number amounts to fewer than 7 tanker loads in 10,000 shipped that were positive for antibiotics by screening assays. Although the rate of antibiotic contamination is quite low, every occurrence is costly to the industry and to the individual producers involved. Milk presumed positive is dumped, time of milk inspectors and insurance adjusters is required to investigate the incident, technical service personnel from antibiotic distributors or manufacturers are asked to explain why their drug "caused a problem" after on-label use, and implicated producers may be fined and not allowed to ship milk for a period of time. The challenge to the dairy industry has been to develop an approach that reduces the incidence of contaminated shipped bulk milk to zero. To achieve this goal, measures need to be developed that can be used by the producer on the farm. One such approach is the development of **HACCP** (Hazard Analysis Critical Control Point) procedures (3, 8).

The objectives of the HACCP program are to identify hazards associated with a process, to determine the acceptable limits in the production process, to determine the critical control points that prevent those hazards, to develop a management strategy that addresses the identified critical control points, to monitor the success of the program, and to make any necessary changes to management procedures in response to the outcome of the monitoring. The HACCP program for antibiotic avoidance, the Milk and Dairy Beef Quality Assurance Program (**MDBQAP**), has been developed jointly by the American Veterinary Medical Association and the National Milk Producers Federation (11). The program defines 10 critical control points for preventing antibiotic residues. These 10 points focus primarily on disease prevention and management of antibiotics and treated cows, although one point is to use screening tests for drug residues.

The producer manual for the MDBQAP suggests that, for antibiotics approved by the FDA Center for Veterinary Medicine for lactating cows, simply following label recommendations on milk withholding times ensures a safe product; however, when antibiotics are used in an extralabel manner, screening tests for drug residues should be used if such tests are available (4). Unfortunately, although the intent of the recommendation is rational, no formal evaluation has been conducted for tests on milk of individual cows (4). The core of the problem is that no regulatory mandate exists for testing milk from individual cows or from the bulk tank, and therefore no incentive exists for kit manufacturers to evaluate their tests under these circumstances. This paper reviews the history of the United States antibiotic testing program up to the development of the current, marketed tests for antibiotic residues, the process used to approve these tests for use on raw, commingled milk, and the current status of the testing program for tanker trucks. Also provided are some epidemiologic guidelines for using the tests validated only for raw, commingled milk to support an on-farm HACCP program for residue avoidance.

DEVELOPMENT OF THE RAPID SCREENING TESTS

Influences Leading to the Development of New Screening Tests

In November 1990, the General Accounting Office (**GAO**) prepared a report to the House Committee on Government Operations (20), stating that the FDA did not have access to appropriate technology to ensure that the national milk supply was free of antibiotics. One of the suggestions in the report was that the FDA should "prioritize and expedite its current efforts to develop and evaluate new screening and confirmatory test methods for animal drug residues in milk. . ." [page 13, (20)]. This report not only stimulated the FDA to consider mechanisms for certifying tests, but also helped to define the need and market for rapid screening tests.

In 1991, the MDBQAP was established as the core of an industry-sponsored national program for residue avoidance in dairy products. As mentioned previously, the program recommends on-farm screening of milk from individual cows as a critical control point. In 1992, the Pasteurized Milk Ordinance (**PMO**) was changed to require all tankers to be tested for β-lactam drugs as they entered milk plants. Although screening tests had been available prior to these events, the GAO report (20), coupled with the changes in the PMO and the interest in on-farm testing, stimulated the development of rapid screening tests by the private sector. By the end of 1992, screening tests were widely available to milk receivers, veterinarians, and dairy producers, but only one of these tests had been evaluated by the FDA and approved as an official test for bulk milk evaluation for β-lactam antibiotics. Few of the tests had any public data on their performance for either in vitro or field trials. In 1992 and 1993, several field evaluations of some screening tests were conducted. The results suggested that the tests were prone to false positives, not only for cows with clinical mastitis but also for cows that were clinically normal (6, 10, 12, 16, 17, 18).

The Need for Evaluation of Residue Screening Tests

Following the public evaluations of the tests, a good deal of criticism and scrutiny was directed at the test manufacturers. The major concern was the lack of rigorous evaluation by the manufacturers, leading to unsupported recommendations for the broad use of the tests. A specific criticism was that the tests were evaluated on augmented milk samples, not on field samples. The manufacturer trials focused on the ability of a test to detect very low antibiotic concentrations, even those below accepted tolerance or action levels of antibiotics, but little or no emphasis was given on the assessment of crossreactivity to natural inhibitory substances in the milk or milk constituents. Also, regard for population or epidemiologic assessment of the tests was minimal.

Despite the limited scope of the evaluations, several manufacturers suggested that the tests could be used to screen for antibiotics in commingled milk from silos, hauling tankers, and bulk tanks as well as milk from cows and individual quarters. Undoubtedly, the need for testing was dictated by public and regulatory pressure, and the dairy industry cannot be blamed for utilizing the available tools, but inadequate evaluation of the tests, coupled with poor epidemiologic interpretation of the test results, led to poor decisions concerning milk disposition from the farm to the corporation.

The real concern over the use of these tests was raised because of the need to evaluate the status of milk of cows on the farm. In the 1980s, herd evaluations (5, 13) of some of the early screening tests indicated that the tests were prone to reporting false positives, but the work of Cullor et al. (9) at the

TABLE 1. The six β-lactam antibiotics and their tolerance or safe levels specified by the FDA to be detected in performance evaluations of antibiotic screening tests.

Antibiotic tested	Tolerance (ppb)
Penicillin G	5
Ceftiofur	50
Cloxacillin	10
Cephapirin	20
Ampicillin	10
Amoxicillin	10

University of California brought industry attention to two serious problems associated with the tests for β-lactam antibiotics: first, that the dairy industry had mistakenly relied on industry claims for the reliability of the tests, and, second, that the dairy industry was asking producers through the MDBQAP to use tests in situations for which they were not designed, that is, to test both bulk tank milk and milk of individual cows. The high rates of false positive results found by Cullor et al. (9) prior to treatment of cows with clinical mastitis cast doubts on whether the tests were applicable for on-farm screening. Subsequent studies (6, 12, 16, 18) verified and extended these results to nonclinical cows for β-lactam screening tests as well as screening tests for tetracyclines and sulfa drugs.

At approximately the same time as the initial paper of Cullor et al. (9), the GAO submitted another report to Congress, accusing the FDA of not having methods for adequate detection of the antibiotics that could potentially be found in the US milk supply (19). The report also recommended that the FDA be more active in moving tests for certifying the safety of the nation's milk supply onto the market.

PERFORMANCE EVALUATION OF THE TESTS

Testing Protocols

Performance evaluation of the tests for raw, commingled milk began in 1992 as the FDA established parameters—or specifications—for evaluation (2). The test specifications focused on four areas. First, specifications described which β-lactam antibiotics were to be detected in augmented samples by the tests as well as the level of detection (parts per billion). Six β-lactam antibiotics were specified, and the detection limits were originally determined as the official FDA tolerance or safe level (Table 1). Official FDA tolerance is the maximum legally allowable level or concentration of a drug in milk or meat at the time it is marketed. Safe levels are used by the FDA as guidelines for deciding whether or not to prosecute, but safe levels are not and cannot be transformed into tolerances. Safe levels are not binding, do not dictate any result, do not limit the FDA's discretion in any way, and do not protect milk producers from enforcement action. Second, the estimated sensitivity of the tests (the proportion of positive samples detected by the screening test) was delineated such that the calculated lower limit of a 95% confidence interval (around the sensitivity estimate) would not be less than 90%. This sensitivity interval had to be attained for at least one of the six specified β-lactam antibiotics. If a test had the ability to detect antibiotics below the FDA tolerance levels with a 90% sensitivity, the manufacturer would be required to state this ability on the label. Because the evaluations were to be conducted on a minimum of 30 augmented milk samples, a successful test would be required to be positive on all 30 samples to meet the interval criteria.

The third area of evaluation for the tests was described as selectivity. The assessment and goals for selectivity mirrored that used for sensitivity and were to be determined using 30 control samples containing no antibiotic. The fourth specification determined the dose-response curve for the tests, which essentially was an evaluation of the limits of detection of the test, i.e., how likely the test was to call a sample positive when the antibiotic concentration in the sample was below tolerance levels. This evaluation was to be conducted on 6 samples of milk that had been fortified with antibiotics over a potential range of 1 ppb to tolerance levels.

Additional experiments were also required to assess "ruggedness of the tests", that is, performance of the tests on frozen samples, crossreactivity of the tests to other antibiotics, interference from bacteria and somatic cells, and stability and consistency among lots. All of these experiments were to be conducted by the manufacturer.

As the FDA was developing the specifications for the test evaluation, they also developed an agreement with AOAC International to supervise the test evaluations, to contract for independent evaluations of the tests, and to certify the tests in compliance with the FDA specifications. Under this agreement, manufacturers submitted their evaluation results and test kits to the AOAC. Evaluation results of the manufacturer were then verified by an independent evaluation laboratory. If a screening test met all aspects of the AOAC evaluation, the test would be certified as performance tested, which would be the stamp of approval from the AOAC and the Center for Veterinary

Medicine. Screening tests that did not meet all the evaluation criteria could still be recognized as valid and certified as performance tested, but acknowledgment of their limitations would be required on the package label.

The Evaluation Process and the Results

The evaluation process officially began in January 1992 with the formation of the AOAC Research Institute. The primary duty of the institute was to administer the test kit confirmation program. The initial period for accepting applications was February 2, 1992 until April 15, 1992. In March, the starting date for application of test review was changed to April 1. In May 1992, the application process was temporarily suspended following the report by Cullor et al. (9) and after voiced concerns by the manufacturers and others about the certification program. Later that month, a meeting was held with manufacturers and others to discuss refinements to the requirements for submission of data. In October 1992, a memorandum of understanding was signed between the FDA and AOAC, which officially recognized the AOAC testing program as an appropriate evaluation of β-lactam tests to be used in state milk monitoring programs. At the time of the signing, the AOAC reopened the application period for test kits. Manufacturers were notified that they would have until January 1993 to submit data for AOAC consideration.

In January, the AOAC reported that applications had been received for the evaluation of 12 test kits. Three were subsequently withdrawn after preliminary review, leaving 9 kits for independent testing and review. In March, the AOAC announced plans to institute a program to evaluate "cowside" tests also. In May, initial results from the testing were distributed to manufacturers, and, after comments and questions, the tests were scheduled for an additional round of testing. Because of additional delays in the certification, evaluation results were not reported until October 1993, at which time the first results of evaluation were reported for 8 tests, and 2 additional tests were reported as certified. By August 1994, 15 tests had been certified as performance tested, and states had begun programs to train personnel in the use of these official tests.

Also in August, the AOAC announced a certification program for screening tests to evaluate raw, commingled bovine milk for tetracyclines, aminoglycosides, and sulfa antibiotics. The Center for Veterinary Medicine also began a limited evaluation of β-lactam screening tests for on-farm use. Also in August, in response to a request from the National Conference of Interstate Milk Shippers, AOAC announced that additional dose-response data would be collected to provide a consistent evaluation for the performance of these tests below tolerance levels for the various antibiotics. In January 1995, two additional tests that were specific for detecting cloxacillin were granted AOAC performance-tested status. As of September 1994, the FDA released the information on 17 β-lactam tests that had been validated in the AOAC-FDA process (14), allowing these kits to be used in the screening and confirmation of the β-lactam status of raw, commingled bovine milk (Table 2).

Consequences of the Performance Evaluation

The history of the evaluations makes it clear that the process of defining attributes of an antibiotic screening and confirmatory test that could please both the kit manufacturers and regulatory officials was difficult. The resulting validation criteria were a compromise, and, therefore, the dairy industry now has a different residue prevention program than originally intended. One intent of the program was to certify rapid tests for regulatory use. The real result of the validation process was a set of tests that detected some of the six β-lactam antibiotics below tolerance levels and others above tolerance levels. With the anticipated regulatory use of these tests, some tankers that have milk containing β-lactam antibiotics at levels below those requiring action will be regarded as tankers shipping milk with illegal concentrations of antibiotics, which puts the milk receivers and those people working in the states charged with enforcing the PMO in a very awkward position.

A second intent of the validation process was to foster commercial development of screening tests that could be used interchangeably in a variety of testing situations and especially to verify results from other tests. The reality is that many accepted tests now exist that detect antibiotics at different concentrations. Results of screening vary among sites, setting up the potential for the shipment of milk that is negative in one state but positive and rejected as contaminated in another state. This concept is inimical to the intent of the PMO to standardize the regulations for milk shipments between states.

A third intent of the validation program was to certify tests for use on raw, commingled bovine milk, not for milk from a cow or even from the bulk tank of

a farm. The reality is that tests carrying the AOAC performance-tested standard will be packaged and marketed as farm tests (Table 2).

A final intent of certification was to develop a battery of tests that could be used across the country to ensure that the milk supply was free of antibiotics. The reality of the certification is that, within the limits of the laboratory evaluations that have been conducted, the tests have been fairly evaluated and validated. The results of these evaluations suggest that, within the matrix of "raw, commingled bovine milk", the tests err on the "safe" side; that is, the tests

TABLE 2. Summary of evaluations of β-lactam screening assays for use under Appendix N of the Pasteurized Milk Ordinance.[1]

Test name	Company	Antibiotic					
		Penicillin G (5)[2]	Ampicillin (10)	Amoxicillin (10)	Cloxacillin (10)	Cephapirin (20)	Ceftiofur (50)
		(ppb)					
Charm II Tablet Competitive Assay™	Charm[3]	4.8[4] (3.5)[5]	9 (6)	10 (7.5)	70 (ND)[6]	4.5 (3)	25 (3.5)
Charm Farm™	Charm	5 (3.5)	10 (4)	10 (5)	40 (ND)	20 (17)	25 (18)
Charm II Tablet Sequential Assay™	Charm	4.8 (3)	8 (4)	10 (6)	50 (ND)	4.5 (3)	23 (7)
Charm II Tablet Transit Test™	Charm	4.8 (2.5)	9 (5)	10 (6)	80 (ND)	4.5 (3)	13 (7)
Charm II Rapid Inhibition Test™	Charm	3 (2.5)	4.5 (2.5)	4.5 (2.5)	25 (ND)	16 (10)	50 (35)
Charm I-Cowside II Tablet™	Charm	4.8 (2.5)	10 (3)	10 (4)	50 (ND)	8 (3)	40 (7)
Charm II Tablet Quantitative Assay™	Charm	4.8 (3)	8 (4)	10 (6)	10 (5)	4.5 (3)	23 (7)
Charm *B. stearothermophilus* Tablet Disk Assay™	Charm	5 (3.8)	6.5 (5)	10 (5.5)	48 (ND)	11 (7)	75 (ND)
Delvo Test® P	Gist-Brocade[7]	3 (1)	10 (2)	8 (2)	30 (ND)	8 (4)	50 (10)
Delvo-X-Press®	Gist-Brocade	5 (1)	10 (2)	10 (1)	50 (ND)	10 (1)	10 (4)
Lactek B-L™	IdeTek[8]	5 (1)	8 (2)	10 (4)	8 (4)	16 (4)	ND (ND)
Lactek CEF™	IdeTek	ND (ND)	ND (ND)	ND (ND)	ND (ND)	ND (ND)	50 (10)
Penzyme III Test®	SKB[9]	5 (1)	10 (4)	8 (1)	80 (ND)	8 (2)	80 (ND)
Penzyme Milk Test®	SKB	5 (1)	10 (2)	8 (1)	80 (ND)	8 (4)	80 (ND)
SNAP Test™	IDEXX[9,10]	5 (3)	10 (6)	10 (6)	50 (ND)	8 (4)	50 (4)

[1]Most evaluations were supervised by AOAC International in accordance to guidelines developed by FDA, AOAC, National Conference of Interstate Milk Shippers, National Milk Producers Federation, kit manufacturers, and American Association of Bovine Practitioners.

[2]The official FDA tolerance or designated safe levels.

[3]Charm Sciences Inc. (Malden, MA). The performance claims of two additional tests (Charm I/Cowside Test for Cloxacillin in Milk and Charm II test for Cloxacillin in Milk) were validated by AOAC in January 1995. The Charm I/Cowside II method was validated at 10 ppb, and the Charm II was validated at 9 ppb both for cloxacillin only.

[4]Values are the validated levels of detection.

[5]Values are the lowest reported level of detection.

[6]Not evaluated for this antibiotic.

[7]Gist-Brocade Food Ingredients Inc. (King of Prussia, PA).

[8]IdeTek Inc. (Sunnyvale, CA).

[9]SmithKline Beecham (West Chester, PA) (now part of Pfizer Inc.).

[10]IDEXX (Portland, ME).

detect the presence of some β-lactam antibiotics below tolerance levels. Because the tests have not been evaluated on a population basis, however, there is a great deal of uncertainty as to how well the tests work in the field on bulk milk from a variety of sources and produced under a myriad of management schemes. There is no epidemiologic context for the application of these tests, and, since January 1995, these tests are undergoing a large, uncontrolled field trial.

Some Preliminary Results of the Field Evaluation

Beginning in January 1995, states were asked to implement screening and confirmatory testing of tanker and producer bulk milk with the validated rapid tests. States implemented the screening and confirmatory testing in different ways, but in essence the programs followed a singular approach. The frequency of testing did not change, and all tanker trucks delivering Grade A bulk milk to processing plants continued to be screened for β-lactam antibiotics. Positive truckloads were then confirmed by a certified laboratory using one of the tests approved by the FDA for confirmation. What changed was the tests being used and the process of screening and confirmation. As previously discussed, the operating parameters of the tests changed so that, in general, the limits of detection moved closer to 0 ppb, and more of the β-lactam antibiotics could be detected. The consequence of these two changes alone would result in an increase in the number of positive tests on bulk tank milk.

Another important change was that screening tests were also being used as the confirmatory test. In all cases, confirmation had used a similar type of test, and, in most cases, confirmation used the same test. This change represented a dramatic departure for most states in the operation of their antibiotic residue program. Previously, many states had relied on an entirely different test for confirmation than that used for screening, in many cases, the *Bacillus stearothermophilus* disc assay. In the best scenario, in which all tests correctly identify contaminated milk, the new rapid screening tests would benefit producer, processor, regulator, and consumer because milk would be quickly and correctly processed. In the worst scenario, in which tests did not correctly identify contaminated milk, a false positive on the screening test (either caused by a nonspecific test or concentrations of antibiotics that were below tolerance) would very likely also occur during the confirmation. Although the ease and speed of testing would continue to benefit the processor and regulator, the potential errors in the testing would be costly to the producer and of no benefit to the consumer. The result of using the same test for both screening and confirmation should be reflected in a very high concurrence of the results between tests as well as an increase in the number of tanker trucks that were reported as contaminated with antibiotics.

Data collected by the Pennsylvania Department of Agriculture for the number of truckloads screened and confirmed positive and the amount of milk dumped are shown for the first calendar quarter of 1994 and 1995 for five states (Table 3). Several important trends may be deduced from this information. First, for those states that have separate screening and confirmation processes (Pennsylvania, Maryland, and New Jersey), the concurrence of screening and confirmation dramatically increased from 1994 to 1995. During the first quarter of 1994, 62% of screened positive truckloads were subsequently confirmed (82 of 132). In 1995, 93% of screened positive loads were confirmed (143 of 154). Second, in the five states, the number of tankers confirmed positive increased 37% between 1994 (n = 174) and 1995 (n = 240). Third, during the first quarter of 1994 in the five states, 2,362,794 kg of milk were dumped. For the same period in 1995, 3,200,208 kg of milk were dumped, a 35% increase from the previous year. The value of dumped milk in the first

TABLE 3. Summary of evaluation of Grade A, raw, commingled milk for the presence of β-lactam antibiotics.

State and year	January to March 1995		
	Screened positive	Confirmed positive	Milk discarded
			(kg)
Pennsylvania			
1994	89	55	769,937
1995	110	102	1,294,198
Maryland			
1994	20	15	193,435
1995	20	19	208,095
New Jersey			
1994	23	12	244,876
1995	24	22	352,006
New York[1]			
1994	. . .	66	710,909
1995	. . .	55	638,182
Ohio[1]			
1994	. . .	26	443,636
1995	. . .	42	707,727

[1]Screening and confirmation are accomplished as a single process.

quarter of 1995 ($12.00/cwt) was nearly $845,000, approximately $221,000 more than the value of discarded milk in 1994. These trends were observed for all states except New York, which underwent little change, probably because New York had had a system of screening and confirmation using the same test in place for some time prior to January 1995. In reality, the numbers shown in Table 2 likely underestimate the true change in incidence of violations because the number of farms in all of the surveyed states decreased.

In Pennsylvania, for the first quarter of 1995, a follow-up was done on 63 of the 110 tanker truck loads that were screened positive. In those cases, the most common test used for confirming a positive screening test was the SNAP Test™ (32 of 63; IdeTek, Portland, ME). The next two most common tests used as confirmatory tests were Lactek (12 of 63; IdeTek Inc., Sunnyvale, CA) and Charm II (12 of 63; Charm Sciences Inc., Malden, MA). In most cases, the screening test and the confirmation test were the same. In these instances there was near perfect agreement in the test results, but when the test used for confirmation was different from the screening test agreement was not as good. In 2 cases, the screening test (SNAP Test™) identified a load as positive that was found to be negative when confirmed using Lactek.

Although these data represent early results of the new testing program, they suggest that dumping of raw milk because of antibiotic contamination has increased over the previous year, because of an increase in the number of loads screened positive and because of the use of the same tests for screening and confirmation. These data support the idea that the increased chemical sensitivity of the tests has had an impact on the program. In addition, additional milk may be dumped because of false positives that were not discovered in the confirmation testing because the same tests were used for screening and confirming.

THE FARM ROLE FOR ANTIBIOTIC RESIDUE TESTS

Reducing the Risk for Antibiotic Contamination of Bulk Milk

In the changing climate of bulk milk testing, it is important that the producer work closely with the veterinarian and field representative to develop an on-farm program to reduce the risk of contaminating bulk milk with antibiotics. Essential components of a risk reduction program include implementing effective programs for disease prevention, developing rational treatment protocols, and maintaining useful records of treatment and disease events. The MDBQAP should be a part of all on-farm programs for risk reduction, because those programs operate on the assumption that testing of milk from treated cows is part of the total program. Unfortunately, no official mandate exists for use of the accepted tests on cow or quarter samples. The AOAC validation is clearly inadequate for tests used on individual cows. The FDA is still in the process of drafting guidelines to evaluate the tests for cow samples, which leaves an obvious void in the efforts of the dairy industry to implement the MDBQAP. Until the void is filled, herd owners and veterinarians must independently assess tests and interpret the results for use on their farm.

One suggested approach is "testing the tests", which provides veterinarians and producers guidelines to evaluate tests on their farms (7). The approach is to evaluate the test on milk from 30 each of untreated cows, treated mastitic cows, and treated mastitic cows beyond the labeled withholding time. This procedure allows a farm-specific evaluation of the tests and, with an understanding of the expected outcomes of such a small-scale trial, the data can be quite useful in making rational choices for on-farm testing (15). Even with this type of protocol, uncertainty regarding the reliability of these tests will probably continue. Ultimately, without more tools and information, the decision of which screening test to use will be difficult at best. The FDA and others are presently evaluating the accepted tests for their suitability for on-farm use. Using this information and results of the "testing the tests" protocol, the industry will have some tools for the on-farm residue reduction program.

Epidemiologic Considerations for Effective Interpretation of Test Results

Regardless of the test used or whether the tests are used on the finished product, on raw, commingled milk, or on milk from an individual cow, the tests must be applied in a manner that maximizes the value of the test results. The usual approach to assessing the performance of a test is to determine the population sensitivity (the probability of a test correctly identifying a cow with antibiotic above tolerance levels in the milk) and population specificity (the probability of a test correctly identifying a cow without antibiotics—or with antibiotics at concentrations below tolerance level—in the milk). It is important to remember the difference between population

sensitivity and the concentration of antibiotic that a test can detect. Similarly, population specificity is more than the ability of a test to differentiate among types of antibiotics. The detection limits of an assay and its ability to distinguish among different chemical moieties are only single factors determining population sensitivity and specificity. More importantly, sensitivity and specificity only partially characterize the usefulness of a test. Predictive value is the most useful measure of the utility of a test.

Positive predictive value is the probability that a positive test result is associated with a cow producing milk containing antibiotics at a concentration requiring action. Negative predictive value is the probability that a negative test result is associated with a cow without antibiotics at a concentration that requires action. Predictive value is a function of population sensitivity and specificity and the prevalence (or probability) of the condition being tested. For a given sensitivity and specificity, predictive values change as prevalence changes. As the probability of a cow being positive for antibiotics increases (i.e., as the prevalence increases), the positive predictive value of the test increases, and negative predictive value decreases. As the probability of a cow being positive for antibiotics decreases, the predictive value positive for the test decreases, and the predictive value negative increases.

The importance of understanding predictive value cannot be understated, because this value helps to guide when a test should be applied. One of the cardinal guidelines for deciding to use a screening test is that the test should provide the practitioner, producers, and regulators with useful information so that they can make rational decisions. To maximize the usefulness, the tests should be applied only to situations in which their predictive values are high. The most appropriate use for antibiotic testing would be to evaluate the negative status of cows following labeled withholding time and the negative status of raw, commingled milk. Use of tests to test cows free of antibiotics prior to the recommended withholding time would be inappropriate, would result in some false negatives, and would jeopardize the milk market of the dairy producer. Even worse would be to use these tests to screen untreated cows in a random manner. The predictive value of a positive test in this circumstance is zero and will only result in the false conclusion that there are cows producing milk that contains antibiotics and is being sold to the consumer. Erroneous conclusions will also be reached when screening commingled milk for which the prevalence of contamination is less than 7 out of 10,000. In this situation, the predictive value of a positive test must be relatively low.

CONCLUSIONS

The validation process for antibiotic residue tests for raw, commingled bovine milk has been arduous. In terms of marketing antibiotic-free milk, the process developed by the FDA and administered by AOAC International can be deemed a success, but the program has created some difficult problems for the insurance industry, processors, producers, and veterinarians. First, the tests have not been evaluated in a population context. The number of samples required for certification was small and was unlikely to represent the range of bulk or individual milk quality that is experienced in the field. The quality of the estimates of the population test parameters resulting from the certification is questionable, which essentially means that the dairy industry has been conducting an uncontrolled field trial since January 1995.

Second, the certification process, which focused on developing convenient tests to be used in the rapid screening and confirmation of milk by processors, allowed tests to be certified that detect some antibiotics below regulatory tolerance levels and, in some cases, above tolerance levels. In addition, because the same tests are being used for both screening and confirmation, there is no check in the system to verify the specificity of the tests. The data from the first 3 mo of 1995 support this contention, and, quite probably, high quality milk is being dumped, and producers are being penalized for inaccurate test results.

Third, some AOAC performance-tested tests or FDA-accepted tests that were intended to be used for raw, commingled milk will be marketed as farm and cow tests. Although the labels for these tests will explicitly describe their approval for commingled milk only, the implicit message from the name of the test is that the test can be used appropriately for individual cow milk, but no data exist to support use of these tests on individual animals, and these tests need to be subjected to protocols evaluating them (7).

Fourth, because the breadth of the testing program is increasing (the previously official test, the *B. stearothermophilus* disk assay detected only a portion of the six β-lactams targeted in the new program), the number of violations detected since January 1995 has increased solely because of the increased ability to detect the antibiotics that had gone undetected previously.

Because these tests are validated and are being actively used, efforts should be undertaken to understand how they work. The goal of the dairy industry needs to be toward the continuing production of milk that is nutritious, good tasting, and safe. Although the tests and their on-farm performance are poorly understood, they still should be utilized as a neces-

sary part of a farm total quality management program and within the context of the MDBQAP.

Tests are not intended to be used to define a quality product, and they should never be used to do so. These screening and confirmatory tests should serve only as the final link in a rational system of antibiotic use that includes on-farm programs for disease prevention, treatment protocols, and effective record-keeping systems. Ultimately, the screening tests used to assess bulk and commingled bulk milk should be viewed as the measure of the success of on-farm programs and not the measure of milk quality. The appropriate system of confirmation testing would focus on eliminating false positives, ensuring that quantitative assays (rather than qualitative assays) that identify the contaminating substance are in place. The consumer and, ultimately, the dairy industry are best served by using tests rationally in support of strong quality programs based on HACCP principles and used to identify problems in the production system rather than using them in a punitive program that has little epidemiologic basis for succeeding or being fair to the producer.

REFERENCES

1 Anonymous. 1995. National Milk Drug Residue Data Base Report, October 1, 1993–September 30, 1994. GLH Inc., Falls Church, VA.

2 AOAC Research Institute. 1992. Specific data submission requirements for test kits to test for beta-lactam residues in milk. AOAC Res. Inst., Arlington, VA.

3 Bauman, H. E. 1974. The HACCP concept and microbiological hazard categories. Food Technol. 28:30.

4 Boeckman, S., and K. R. Carlson. 1994. Page 25 *in* Milk and Dairy Beef Residue Prevention Protocol. 1995 Producer Manual. Agri-Education Inc., Stratford, IA.

5 Carlsson, A., and L. Bjorck. 1987. The effect of some indigenous antibacterial factors in milk on the growth of *Bacillus stearothermophilus* var *calidolactis*. Milchwissenschaft 42:282.

6 Carlsson, A., and L. Bjorck. 1992. Liquid chromatography verification of tetracycline residues in milk and influence of milk fat lipolysis on the detection of antibiotic residues by microbial assays and the Charm II test. J. Food Prot. 55:374.

7 Cullor, J. S. 1994. Testing the tests intended to detect antibiotic residues in milk. Vet. Med. 89:462.

8 Cullor, J. S. 1995. Implementing the HACCP program on your client's dairies. Vet. Med. 90:290.

9 Cullor, J. S., A. Van Eenennaam, J. Delinger, L. Perani, W. Smith, and L. Jensen. 1992. Antibiotic residue assays: can they be used to test milk from individual cows? Vet. Med. 87:477

10 Cullor, J. S., A. L. Van Eenennaam, I. A. Gardner, L. Perani, J. Dellinger, W. L. Smith, T. Thompson, M. A. Payne, L. Jensen, and W. M. Guterbock. 1994. Performance of various tests used to screen antibiotic residues in milk samples from individual animals. JAOAC (J. Assoc. Offic. Anal. Chem.) 77:862.

11 Herrick, J. B. 1991. Milk quality assurance and dairy practitioners. JAVMA (J. Am. Vet. Med. Assoc.) 199:1268.

12 Hoffmeister, A., G. Suhren, H. Braun, and W. Heeschen. 1992. Sulfonamidnachweis in Milch mit Mikrobiologischen, Immunologischen und Chemischphysikalischen Methoden. Dtsch. Milchwirtsch. 43:724.

13 Jones, G. M., and E. H. Seymour. 1988. Cowside antibiotic residue testing. J. Dairy Sci. 71:1691.

14 Milk Safety Branch, editor. 1994. Beta-lactam testing under Appendix N of the PMO (FDA Memorandum M-a-85). Dep. Health Human Serv., Publ. Health Serv., FDA CFSAN:OFP: DCP:Milk Safety Branch, Washington, DC.

15 Sischo, W. M. 1995. Using residue tests in support of a quality milk program. Proc. Am. Assoc. Bovine Pract. 27:40.

16 Sischo, W. M., and C. M. Burns. 1993. Field trial of four cowside antibiotic-residue screening tests. JAVMA (J. Am. Vet. Med. Assoc.) 202:1249.

17 Tyler, J. W., J. S. Cullor, R. J. Erskine, W. L. Smith, J. Dellinger, and K. McClure. 1992. Milk antimicrobial drug residue assay results in cattle with experimental, endotoxin-induced mastitis. JAVMA (J. Am. Vet. Med. Assoc.) 201:1378.

18 Van Eenennaam, A. L., J. S. Cullor, L. Perani, I. A. Gardner, W. L. Smith, J. Dellinger, W. M. Guterbock, and L. Jensen. 1993. Evaluation of milk antibiotic residue screening test in cattle with naturally occurring clinical mastitis. J. Dairy Sci. 76:3041.

19 Zadjura E. M., W. M. Layden, M. B. Karpman, P. Turner, J. F. Mitchell, J. Rose, K. Van Egmond, P. Ward, and L. Moore. 1992. Animal drug residues in milk. United States General Accounting Office, Washington, DC.

20 Zadjura E. M., M. J. Rahl, M. L. Aguilar, W. M. Layden, S. W. Weaver, and M. B. Karpman. 1990. Page 13 *in* Drug Residues in Milk. United States General Accounting Office, Washington, DC.